DOCUMENT RESUME

ED 312 749 EA 021 396

AUTHOR Haertel, Geneva D.; And Others
TITLE Capturing the Quality of Schools: Approaches to Evaluation.
PUB DATE Mar 89
NOTE 52p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Accountability; Accreditation (Institutions); Educational Assessment; *Educational Quality; Elementary Secondary Education; *Evaluation Methods; Excellence in Education; *School Effectiveness

ABSTRACT

This document reviews several approaches used to examine schools, evaluate their quality, or compare them to one another. The rationale and major purposes of each approach, the variables and processes employed, and the potential contributions of that approach to a comprehensive evaluation model are addressed. Six approaches are covered: (1) models used in state-level accountability systems; (2) models used in school recognition programs; (3) effective schools research paradigm; (4) self-study approaches; (5) models used in the accreditation process; and (6) models based on rich, contextualized descriptions of schools. The various approaches focus primarily on either school-process variables or outcome variables; few implementations offer thorough coverage of both. Drawing on the discussions of these six approaches, this paper then presents some implications for a methodology of comprehensive school evaluation. Examples of variables/indicators for use in comprehensive school-level evaluation, and two figures are appended. (SI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED312749

Capturing the Quality of Schools: Approaches to Evaluation

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Geneva D. Haertel
Independent Consultant
Stanford, California

Conrad G. Katzenmeyer
Office of Educational Research and Improvement[1]
United States Department of Education
Washington, DC

Edward H. Haertel
School of Education
Stanford University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Geneva D Haertel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: SCHOOL EVALUATION

Paper presented at the meeting of the American Educational Research Association, San Francisco, California, March 1989.

[1]This manuscript was coauthored by Conrad G. Katzenmeyer in his private capacity. No official support or endorsement by the United States Department of Education is intended or should be inferred.

EA 021 396

BEST COPY AVAILABLE

Capturing the Quality of Schools: Approaches to Evaluation

Introduction

Educational curricula, programs, and policies are implemented in a complex school context. The self-contained classroom may be the immediate locus of most learning activities, but researchers have come to recognize that school-level factors can be critical in determining learning outcomes (e.g., Berman & McLaughlin, 1977; Purkey & Smith, 1983). Concerns with the quality of education are increasingly cast in terms of the quality of schools. The United States Department of Education (USDE) recently established a Center at Stanford University to examine the secondary schools as contexts for teachers' work. USDE also has established national school recognition programs at the elementary and secondary level. California, South Carolina, as well as other states have established programs for recognizing exemplary or distinguished schools. President Bush has requested $500 million to identify and reward exemplary schools across the nation. Several months ago, an invitational conference was held at the White House to promote programs under which parents may choose the schools their children may attend. Parents in Minnesota and elsewhere have greater freedom than ever before to decide which school their children will attend, and with that freedom has come an increased need for trustworthy, relevant information in a form that is useful for comparing one school to another. For these reasons, the evaluation of schools has taken on increased importance.

A comprehensive school evaluation should (1) describe a school and (2) diagnose problems in its functioning. In describing a school, the

evaluation should portray the community and student population it serves; the school's distinctive goals, strengths, and limitations; and the programs and processes through which it strives to meet both its own particular goals and the common goals of the school system. The description should also indicate outcomes for students, teachers, parents, and perhaps even the community, and should provide some basis for determining the quality of those outcomes. Norm-referenced comparisons of test score means among schools are just one limited example.

In order to diagnose any problems in the school's functioning and suggest possible avenues and approaches for improvement, a comprehensive evaluation must examine the school's instructional processes and approaches, as well as its formal and informal administrative and decision making processes. It must elicit the views of the school's participants concerning problems and solutions, as well as inferring problems from evidence of poor learning outcomes, negative parent or community sentiment, or other sources. Above all, these different forms of information and evidence must be brought together to create a coherent picture, a sort of causal model, of the school's functioning. Mere description will not suffice.

Clearly, the kind of school evaluation envisioned differs substantially from the evaluation of curricula or programs within schools. To date, the methodology of program evaluation is better developed and more widely understood that school evaluation. There is no single, accepted model for a comprehensive school evaluation, but various systems and approaches have been developed for describing particular schools, ranking groups of schools, or judging schools against established standards.

This paper briefly reviews several approaches now used to study schools, evaluate their quality, or compare them to one another. It addresses the rationale and major purposes of each approach, the variables and processes employed, and the potential contributions of that approach to a comprehensive evaluation model. Six approaches are covered, including (1) models used in state level accountability systems, (2) models used in school recognition programs, (3) the effective schools research paradigm, (4) self-study approaches, (5) models used in the accreditation process (e.g., North Central), and (6) models based on rich, contextualized descriptions of schools.

The six approaches reviewed differ in important ways. Without intending any rigid "matrix" organization, and without minimizing the substantial variability among implementations of any one approach, Figure 1 depicts some of these differences. As indicated by the Figure, the various approaches focus primarily on either school process variables or outcome variables; few implementations offer thorough coverage of both. The vertical dimension of the figure depicts the comprehensiveness, thoroughness, or "richness" of the models. In general, greater investments of time and resources yield greater returns of information from a school evaluation.

Insert Figure 1 about here

Drawing on the discussions of these six approaches, the paper then presents some implications for a methodology of comprehensive school evaluation. A model evaluation would include input, process, and outcome

variables, and would draw these variables together into a coherent picture of the school's functioning, providing both description and diagnosis.

Although the proposed evaluation methodology might be used with private or sectarian schools, the focus in this paper is primarily on the public school sector. Private schools differ substantially from one another with respect to goals, sources and levels of fiscal and other resources, clientele, curriculum, and instruction. As a group, private schools also differ sharply from public schools. Limiting the discussion primarily to public schools will not change the nature of the analyses and critiques presented, but will obviate the need for many cavaets and exceptions.

Educational Accountability (Indicator) Systems

In recent years, state governments have assumed increasing responsibility for monitoring and improving the educational system. Public attention has been focused on the quality of educational outcomes, creating pressures for state legislative action; and patterns of school finance have shifted so that the proportion of funding from the state level has increased. In response to such pressures, nearly all states have implemented accountability systems of one kind or another (Council of Chief State School Officers, 1987). These accountability systems typically involve the collection, organization, and reporting of school-level variables from various sources. State-level targets may be specified for the different indicators or, more typically, school-level norms of some kind may be prepared. The most meaningful interpretive information is likely to come from comparisons with a school's own performance in prior years. School "profiles" or "report cards" are

prepared, featuring state indicators with associated targets, rankings, or longitudinal comparisons. These reports often provide for the addition of locally developed indicators, as well.

Most state accountability systems have relied primarily on data already obtained for other purposes. State education agencies have historically collected several categories of information about schools and school districts, including data on educational finance, enrollments, staff size and credentials, and usually student achievement (OERI State Accountability Study Group, 1988). Federal reporting requirements for categorical programs administered through the states have also led to the collection of data on students eligible for and enrolled in bilingual, Chapter 1, and other compensatory education programs, as well as counts of children with specific handicapping conditions. Because these various data collection activities have been initiated for different purposes and administered under different auspices within the state bureaucracy, however, there has often been little coordination of data collection or integration of the information collected. The development of educational accountability systems has led to some consolidation of state education data, but these data have not been adequate to create systematic, coherent, and comprehensive indicator systems. The variables included may be informative in themselves, but taken together they often fail to address important aspects of school structure, function, and outcomes (OERI State Accountability Study Group, 1988).

State testing, accountability, and educational indicator systems, and similar systems at the federal and local levels, differ considerably in their design, and may be used in different ways to influence schooling

processes and outcomes. One of the most common mechanisms to influence educational practice is simply to publicize performance data. When schools' rankings with respect to mean test scores appear in local newspapers, they generally attract considerable attention. Principals and even teachers in low-performing schools may experience substantial pressure to improve from both school boards and the public at large. Because they are often used by realtors as indicators of educational quality in different communities, a school's test scores may even influence the surrounding community's property values.

Performance data may be used in other ways to influence educational processes. In some states, school accountability systems determine the allocation of rewards to high-performing schools. These rewards usually take the form of some public recognition, but may include extra resources or waivers of specified regulations or requirements. Accountability data may also be used to identify schools or districts in need of technical assistance. In a few cases, serious deficiencies revealed by state monitoring systems may trigger strong, direct state intervention in the operations of school districts (OERI State Accountability Study Group, 1988).

Variables Included

Variables representing educational outcomes are prominent in all state accountability systems. Nearly all such systems feature student test scores, collected using instruments labeled either achievement tests or competency tests. These are often limited to multiple-choice exercises, although writing samples are becoming increasingly popular. Scholastic Aptitude Test (SAT) or American College Test (ACT) scores may

also be used, although their interpretation is complicated by fact that groups of students taking these tests are self-selected. Attendance, dropout, and graduation rates are also widely used; as are fiscal and administrative data including staffing patterns, teacher credentials, pupil-teacher ratios, and per pupil expenditures (Oakes, 1986).

Data on educational processes are less accessible for state accountability purposes, although course taking data may be reported, including enrollments in Advanced Placement (AP) courses, foreign languages, science, or advanced mathematics courses, as well as art, music, and other special subjects. Students may also be asked to report on the number of writing or homework assignments they receive, leisure reading, television viewing, and other behavior.

Because educational outcomes are strongly influenced by some factors beyond a school's control, nearly all states collect background data of some kind, and about half of the states use such data to help guide their reporting and use of outcome variables. For example, schools' achievement score means may be reported in the context of data on student mobility, race/ethnicity, or language background; or parental income and education.

Any of several methods may be used for incorporating background information into test score reporting. In one approach, schools are stratified according to a socioeconomic composite, and achievement levels are compared within strata. A closely related method employs "floating" comparison bands, in which schools are ranked according to a socioeconomic composite and each school's achievement means are then compared to those in its own reference group, consisting of schools within a fixed number of ranks above and below it. In another approach,

achievement is regressed on background variables; predicted achievement levels are calculated for each school; and a school's actual achievement is then compared to its predicted achievement. In at least one state, cluster analysis was used to define distinct community types. School comparisons are made within community types, with a regression adjustment for socioeconomic level (Haertel, 1989; OERI State Accountability Study Group, 1988).

Comparisons of outcomes among schools facing different degrees of educative difficulty are inherently unfair, and even though adjustments based on stratification, clustering, or regression approaches may reduce the degree of this unfairness, they are unlikely to eliminate it. Moreover, with such adjustments there is a risk of legitimating present inequities, of fostering the belief that a school doing as well as others serving similar students is doing well enough (Haertel, 1989).

Educational Accountability Systems and School Evaluation

There are several difficulties with educational accountability or indicator systems as models for comprehensive school-level evaluation. These systems employ a limited range of variables, and tend to focus excessively on outcome measures, especially objective test scores, rather than measures of educational processes. Most state-level systems do not provide any coherent model of a school's functioning, and so are of limited value for either describing schools or diagnosing their problems. Finally, although some of the more successful accountability systems are designed and operated by local school districts, those managed from the state level are unlikely to promote the kind of systematic, cooperative effort among teachers and principals that is needed to effect significant school change.

The variables employed in educational accountability sytems, especially those at the state level, tend to be assembled largely from data already collected for different purposes. These collections of variables sometimes fall short of the kind of coherent, integrated models of schooling required for adequate school-level description or diagnosis. They are quite limited in their capability to diagnose the causes of poor schooling outcomes or to indicate approaches to school improvement. Without such diagnostics, there is little that low-performing schools can do about their poor standing, except to try to influence tested outcomes directly. They may increase the time devoted to tested outcomes, and may even offer drill and practice on items similar to those tested. In addition to constricting both curriculum and instruction, this narrow focus may compromise the validity of the tests as measures of even a limited range of learning outcomes.

A study by the Center for Policy Research in Education (CPRE) between 1985 and 1987 confirmed the reality of these concerns. Roughly 350 policy makers, teachers, and school personnel in four states were interviewed about the impact of accountability systems on schools and classrooms. It was found that accountability systems could indeed influence local educational planning processes and teaching activities, but their effect was often to focus instructional activities narrowly on the indicators, and not to effect any more general school improvement. The accountability system in Minnesota, which is locally designed and operated, was found to be more successful than centralized systems in other states in promoting broader educational change and improvement (OERI State Accountability Study Group, 1988).

An analysis of local indicator systems conducted for CPRE by Jane David (1987) may help to explain the greater success of the Minnesota system. David argues that if indicators are to be used at the local level to promote goals set by districts, then local schools and districts must identify the indicators to be used. She suggests that a system of indicators will be useful in guiding school improvement if it includes measures of the content and quality of instruction, and if the analyses and presentation of data bear directly on specific policy issues. Even valid, locally developed indicators will not be sufficient in themselves to catalyze change; David also identifies five organizational factors that may encourage the use of data for school improvement: a supportive organizational climate; commitment to improvement on the part of district leaders; stakeholder participation in selecting the indicators; technical support for analyzing and reporting data; and development of the system's capability to initiate and sustain change.

Despite their limitations, indicator and accountability systems do possess some features that would contribute positively to a comprehensive school evaluation model. At their best, indicators can provide benchmarks for measuring educational progress (e.g., higher achievement scores, lower dropout rates, or fewer violations of school rules), and can represent aspects of educational process that are plausibly related to educational outcomes (e.g., instructional time). They can capture key descriptors of the educational system (e.g., curriculum offerings, teacher work load, or fiscal information), direct attention to present or potential problems, and inform policy decisions. The most useful indicators for these purposes will be valid and reliable, readily

interpretable, inexpensive to collect, and of enduring significance. If indicators are to be compared across schools, they should also be broadly relevant, and should be defined uniformly across schools (Oakes, 1986).

School Recognition Programs

School recognition programs are designed to identify and publicize unusually successful schools, on the assumption that friendly competition may stimulate better school spirit and improved outcomes among schools in general. Under most programs, those schools that satisfy eligibility criteria and choose to participate must prepare a fairly lengthy written application to the agency sponsoring the program. Panel reviews of applications follow, and site visits are conducted for finalists. Currently, such programs are sponsored or managed by the federal government, states, universities, and private industry (Wynne, 1988).

Recognition programs represent another approach to using public recognition to influence educational practices. These programs are free of some potential disadvantages shared by school accountability programs, in that most are entirely voluntary and all are intended to provide only positive rather than adverse publicity. The identities of schools that lose in competition are not released (Wynne, 1988). Of course, these programs may generate discord despite their voluntary nature and positive focus. If there are only two middle schools in a district, for example, recognition of either could have invidious consequences, whether or not the other had also met eligibility criteria and elected to apply.

One of the earliest school recognition programs was established by the Ford Foundation in 1982. In 1983, that program was joined by one under federal auspices to identify exemplary high schools, and there are

now federal school recognition programs at both the elementary and secondary levels. Since 1984, California, South Carolina, and Florida have established state-level recognition programs, and there is also a school recognition program housed at the University of Illinois at Chicago (Peterson, 1988).

Variables Included

Most state-sponsored programs cover both elementary and secondary schools. They vary considerably in their application processes, their recognition criteria, the value and form of the rewards offered, and the freedom they give local districts to define award criteria and select recipients (Peterson, 1988). Some programs, like South Carolina's, are entirely automatic. Schools are screened on student achievement gain, student attendance, and teacher attendance, and those meeting the achievement gain criterion receive a fixed monetary award per pupil, as well as school incentive reward flags and certificates. If one or both attendance criteria are met in addition, the monetary award is larger. Roughly a quarter of South Carolina's schools qualify for some award in a given year (May, 1987).

California's School Recognition Program begins with an automatic screening on achievement and other performance indicators from the state's school profiles. Based on this screening, Outstanding Achievement awards are provided automatically for several indicators, and schools showing a pattern of exceptional performance are nominated for recognition as California Distinguished Schools. Nominated schools are invited to complete an extensive written application describing their various programs and accomplishments, and site visits to all those making

application are conducted by state and county education representatives (California State Department of Education, 1986).

In Florida, the state encourages local school districts to establish recognition programs, with the cooperation of teachers unions. Dade County's program, for example, requires that teachers as well as principals vote on participation in the school recognition program, and gives individual employees the right not to participate, regardless of the school's decision. Recognition is based on achievement test score gains; level of participation on a standardized physical fitness test; and, for the higher of the two award categories, a plan developed at the school level to correct or improve some aspect of student achievement. Winners receive monetary awards in a fixed amount per participating school employee (Dade County Public Schools and United Teachers of Dade, 1986).

The United States Department of Education (USDE) National Elementary School Recognition Program combines some features of several of the state programs. To be eligible, an elementary school must have at least three grade levels and its own administrator, and must satisfy either of two criteria for achievement in reading and mathematics. In general, schools are eligible if at least 75 percent of their students are at or above grade level in each content area, or if they have shown a pattern of steady improvement over the past three years and presently have at least 50 percent of their students at grade level. The tests to be used and the definition of "grade level" are left to the discretion of each state or the local school district. Those schools meeting the eligibility criteria and electing to participate must complete applications documenting the quality of the school organization, building leadership, curriculum and

instructional program, classroom instruction, school climate, school-community relations, efforts to maintain quality and improve, and student outcomes. They must strive to develop character as well as promoting learning. Panel reviews of applications and site visits follow (Peterson, 1988).

There is a tension in school recognition programs between explicit, objective selection criteria and more flexible, subjective criteria. Explicit criteria may encourage an unhealthy standardization of educational programs and approaches, such as a narrow focus on improving scores on standardized tests. In this way, they may penalize schools with different goals or different strengths. On the other hand, more subjective criteria may be unreliable and difficult to administer. If each school is permitted to prepare its own description supporting the quality of its instruction, school climate, school/community relations, or other indicator dimensions, then the selection of winning applicants may be unduly influenced by the personal preferences of the judges, or the literary skill of the school staff or other writers who prepare the narrative descriptions.

School Recognition Programs and School Evaluation

School recognition programs are vulnerable to all of the problems inherent in school rankings and comparisons. Rankings dependent on measured school performance are inherently biased toward schools in more affluent neighborhoods, and adjustments based on stratification or regression methods are at best imperfect. More affluent schools may also be able to devote greater resources to preparing their applications.

Another difficulty with the use of objective test scores for selecting award recipients is the instability of school-level score rankings across grade levels, content areas tested, and over time. Of course, raw score rankings of schools are quite stable across these dimensions, but that stability is largely associated with differences in the student populations different schools serve. Scores adjusted to remove variation associated with socioeconomic differences are much less stable. Mandeville (1988; Mandeville & Anderson, 1987) has found that even if a given school ranks highly in its adjusted third grade reading scores, say, for several years running, it may well be no more than average at a the second or fourth grade levels. He concluded that grade-within-school effects dominated global school effects at the elementary school level.

The voluntary nature of school recognition programs would also be problematical in a comprehensive school evaluation model. Because they are designed to identify and reward excellence, these systems are necessarily insensitive to problems and difficulties. Schools in trouble are unlikely to volunteer for scrutiny they can just as well avoid. Even if problems were identified, this evaluation approach would be unlikely to generate recommendations for improvement.

Effective Schools Research Paradigm

Beginning in the early 1970s, a new paradigm emerged in the search for effective educational approaches. Large-scale studies, notably Coleman, et al. (1966), had found few or no school-level variables consistently related to average learning outcomes. In response to such discouraging findings, Dyer (1972) suggested a different way of using regression analyses to identify effective schooling practices. Rather than

examining regression coefficients, regression residuals would be studied to find particular schools where learning outcomes exceeded the levels predicted from a regression of achievement test scores on socioeconomic factors. Unusually effective schools, identified by their large positive residuals, could then be studied more intensively to discover the keys to their success. Klitgaard and Hall (1973) applied this approach to six data sets, and other studies followed.

Variables Included

Out of these studies, a loose consensus emerged on the characteristics of "effective schools" (Austin, 1979), and over the next several years, attempts to capitalize on effective schools research by "implementing" these variables took on some of the character of an educational movement. The somewhat disparate findings of early studies were distilled into a simple recipe for school improvement (Purkey and Smith, 1983), and a five-factor "model" for effective schools emerged, described by Ralph and Fennessey (1983, p. 694) as including "some combination of: 1) strong administrative leadership, 2) a safe and orderly school climate, 3) an emphasis on basic academic skills, 4) high teacher expectations for all students, and 5) a system for monitoring and assessing pupil performance." Aspects of effective schools that were more difficult to implement or assess, such as teacher flexibility and positive classroom climate, received less attention over time, and effectiveness came to be identified with a narrow range of tested outcomes. Cuban (1983, p. 695) described effectiveness as a constricted concept, "tied narrowly to test results in mostly low-level skills in math and reading," and "[ignoring] many skills, habits, and attitudes beyond the reach of paper-and-pencil tests."

The Effective Schools Model and School Evaluation

The effective schools perspective is more a rhetoric of reform than a scientific evaluation model (Ralph & Fennessey, 1983). Lists of effective schools characteristics reported by different investigators are not entirely consistent, and their empirical base is weak. In many effective schools studies, student background characteristics were poorly controlled, so that even the identification of some schools as particularly effective may be in doubt. Moreover, measurements and observations used to contrast more and less effective schools were sometimes of questionable reliability and validity. Writing on personnel evaluation, Scriven (1987) questioned the validity of measuring effectiveness using variables correlated with effectiveness, but not directly measuring effectiveness. The same criticism may be leveled at many of the variables included in effective schools models.

Even if effective schools could be identified unambiguously and their distinctive features could be determined, it would not necessarily follow that other schools could become more effective by emulating those features. The so-called five-factor model based on Edmonds' (1979) review is far from a comprehensive and coherent model for schooling processes, and more important, the effective schools literature's specific implications for action are unclear (Cuban, 1983). Ineffectual principals cannot become strong leaders by a simple act of will, nor can teachers change their expectations overnight. School climate is a complex and subtle concept, difficult even to define, and resistant to change by administrative fiat.

Despite these shortcomings, the basic tenets of the effective schools movement continue to enjoy popular support. It may be possible to capitalize on effective schools concepts in fashioning a comprehensive school evaluation model. In their critical review of the effective schools literature, Purkey and Smith (1983) derive a somewhat speculative portrait of the culture of an academically effective school, describing its structure, its process, and a climate of values and norms that emphasizes successful teaching and learning. Their conception is consistent with the popular image of an effective school, but is much closer to the kind of coherent model required for a comprehensive school evaluation.

Based on their review, Purkey and Smith (1983) suggest a set of nine organization-structure variables and four process variables that goes beyond the five "effective schools factors" in its implications for action that a school might take. They go on to suggest a strategy for change consistent with the view of schools as "loosely coupled systems" (Meyer & Rowan, 1978) and with research on the implementation of educational change (Berman & McLaughlin, 1977; McLaughlin, 1978). The organization-structure variables Purkey and Smith propose include (1) school-site management, (2) instructional leadership, (3) staff stability, (4) curriculum articulation and organization, (5) schoolwide staff development, (6) parental involvement and support, (7) schoolwide recognition of academic success, (8) maximized learning time, and (9) district support. Their four process variables are (1) collaborative planning and collegial relationships, (2) sense of community, (3) clear goals and high expectations, and (4) order and discipline.

Self-Study Approaches to School Improvement

The effective schools movement has led to the development of various packaged systems and other resources designed to help schools become more effective by implementing its precepts. For example, Research for Better Schools (RBS), in conjunction with the New Jersey School Boards Association, has developed a set of materials called "Sizing Up Your School System" to guide districts through a self-study process based on the effective schools concepts (Buttram, Corcoran, & Hansen, 1986). RBS offers technical support to assist school boards in identifying standards, choosing or developing instrumentation, data collection and analysis, and preparation of reports on the findings of the self-study.

Another such resource, briefly described below, is a book by Edward F. DeRoche (1987) that is intended to assist administrators in conducting a comprehensive school evaluation and initiating constructive change. DeRoche begins with a review of the effective schools literature, summarizing lists of critical factors or features proposed by several different authors. It emphasizes the importance of a team effort in both evaluation and school change, and offers concrete methods to assure broad participation by teachers especially, but also by students, their parents, and the public.

Variables Included

DeRoche's book features over 75 ready-to-use forms and instruments, most designed for use by teachers. Most of these elicit opinions and perceptions, rather than objective information, and are used as a stimulus to discussion and participation, and as a point of departure for planning improvements. Recommendations for data analysis are limited for the

most part to tabulations and histograms of responses. Little information is presented in DeRoche's book concerning the reliability and validity of these instruments, nor is such information required in a book of this kind, but extensive source citations and other references are provided.

DeRoche's discussion and instrumentation are presented in chapters on the evaluation of the school culture and classroom climate; the principal's instructional leadership and supervision; classroom instruction; and the curriculum. Additional chapters address areas that are less central to the effective schools model, including the effectiveness of the student activities program; pupil personnel services and personnel; school-community relations; office, food, and transportation services; and the management of the school plant and facilities. Specific subareas are discussed in each chapter.

Self-Study Approaches and School Evaluation

DeRoche emphasizes the importance of a locally based evaluation, and of consensus and participation on the part of the school staff. He recognizes that schools need to respond to state accountability or evaluation systems, but suggests that useful local evaluations will be considerably broader than most state-level systems. DeRoche offers many practical activities and suggestions that appear likely to support a positive, constructive evaluation process.

Self-study systems place the initiative for change and improvement squarely on the shoulders of the school administration and faculty. The methods proposed could easily be implemented in a superficial fashion that would not lead to any authentic improvement at all. On the other hand, serious, long-term, systematic self-study appears to be among the

more promising approaches to school improvement, and "packaged" systems may aid in its implementation.

Accreditation Models

School accreditation began in an era of much greater heterogeneity among educational institutions. At one time, it served as a selective, discriminatory mechanism to assure elite colleges that graduates of certain high schools had been exposed to a rigorous course of study. Over time, it has become less selective and more formative in character. Today, it serves as a quality control mechanism by helping to assure conformity to accepted standards in the delivery of educational services, and by encouraging reflection and self-study by a school's faculty and administration (Bryant, 1986). Accreditation models are most fully developed at the secondary school level.

In most states, one of six regional associations or agencies control the accreditation process, and accreditation is formally a conferral of membership in that organization (Mayhew, 1982). School membership is voluntary throughout all regions. Once approved for accreditation, a school peridically undertakes self-study using instruments developed by its regional association. Most schools are granted membership for the next six years, with shorter terms generally being regarded as sanctions. The accreditation process itself consists primarily of self-study by the institution to be accredited; one or more visits by an external examining committee; and formal documentation of the school's strengths and weaknesses, the examining committee's recommendations, and the committee's decision concerning the level and duration of accreditation to be conferred (Bryant, 1986).

School accreditation is concerned almost exclusively with educational inputs and processes, rather than outcomes. The accreditation process rarely involves any use of test scores or other quantitative performance indicators. It is driven by a set of standards to be met, concerned largely with the school's facilities and other resources, written policies, administration and staffing, and curriculum. If an institution falls short on one or more of these criteria, the remedy is usually clear. The self study required as part of the accreditation process may in principle be the centerpiece of a thorough formative evaluation, but if the self-study is limited to satisfying the letter of the accreditation standards, it is unlikely to uncover problems of which a school was unaware.

Variables Included

Viewing the accreditation process from the perspective of school evaluation, the variables to be measured are embodied in the accreditation standards, and the methods of measuring those variables are represented in the accreditation procedures. Historically, accreditation standards have addressed such matters as the scope of the curriculum offered; the number of teachers, their degrees, teaching credentials, and compensation; teaching loads and pupil-teacher ratios; physical plant, including library and laboratory facilities; policies concerning staff development; and records of attendance and pupil progress. Standards concerning actual instructional processes or school climate (e.g., academic focus) were generally evaluated in terms of stated philosophies or reports gathered through interviews by the site visitation team (Bryant, 1986).

Modern accreditation standards typically cover the same general areas, but the precise definition of variables is driven to a larger extent

by a school's own definition of its purposes and functions. The processes of accreditation serve more to promote careful self-study than to gather uniform evidence about variables that are rigorously defined and carefully standardized. For example, the criteria for accreditation by the Western Association of Schools and Colleges (WASC) begin with a requirement for a statement of a school's philosophy, goals, and objectives, developed through a process involving the community, administration, staff, students, and governing board. The WASC standards then go on to discuss requirements for a school organization, student personnel services, curricular program, co-curricular program, staff, school plant and physical facilities, and financial support that are aligned with the statement of goals and philosophy (Accrediting Commission for Schools, 1981). Self-study is encouraged by the process of formulating the initial statement and also by the forms of documentation required in specific areas, but the evidence obtained in many categories is unlikely to be directly comparable across schools.

Accreditation and School Evaluation

The principal weaknesses of modern school accreditation as a comprehensive evaluation model are its limited reliance on objective data, lack of attention to outcomes, and heavy dependence on self-reports by the faculty and administration of the school evaluated. In addition, members of site visitation teams tend to be drawn from the ranks of school administrators, who may be more sympathetic to standard practices than to bold departures (Bryant, 1986). Scriven (1972) has argued that accreditation has become a largely symbolic process, serving to legitimate those schools which conform to accepted organizational

patterns and activities. Meyer and Rowan (1981, p. 81) concur, describing education as "a certified teacher teaching a standardized curricular topic to a registered student in an accredited school." Given the fallibility and limited range of available schooling outcome measures, however, and our limited knowledge of relationships between input and process variables, there is something to be said for a direct examination of educational processes and for the enforcement of normative standards.

Notwithstanding these potential criticisms, some elements of the school accreditation process might be incorporated into a comprehensive school evaluation model. The process of self-study initiated in accreditation seems healthy. Research by Harkins (1981), Gatley (1975), and Telford (1976) affirms educators' belief that self-study can engender school improvement. In a recent doctoral dissertation, however, Bryant (1986) was unable to locate any empirical evidence that self-study actually led to improvements in student achievement test scores or other quantifiable learning outcomes. Preparing statements of philosophy, goals and policies; guidelines for staffing, teaching, and curriculum; and other components of accreditation may also contribute to a school's effectiveness by encouraging a dialogue among school personnel.

Descriptive Studies

All of the foregoing models and approaches have been designed for routine use with large numbers of schools. In contrast to such large-scale systems, there is also a rich case study literature in education, which uses narrative descriptions of a few selected schools to illuminate the character and complexity of all schools. Many of these studies draw on an increasingly diverse and sophisticated range of naturalistic methods,

including different forms of ethnography, naturalistic inquiry (Guba, 1987) and educational connoisseurship and criticism (Eisner, 1983). Others (e.g., Goodlad, 1984) rely more heavily on objective measures and numerically quantifiable data to fashion their descriptions.

Recent works that could be considered naturalistic descriptive studies include Horace's Compromise (Sizer, 1984), The Good High School (Lightfoot, 1983), and The World We Created at Hamilton High (Grant, 1988). A Study of Schools, reported in A Place Called School (Goodlad, 1984) and in other books and articles, illustrates the use of more quantitative data in combination with naturalistic methods and observations. Regardless of the methods used, descriptive studies attempt to convey a comprehensive, integrated description of schools and schooling. At their best, they display the diversity among schools and the contexts in which they function, and the many perspectives and perceptions of students, teachers, administrators, parents, and other stakeholders. These descriptive studies have gone far beyond superficial rankings of schools or evaluations against lists of standards, and have created evocative and illuminating portraits of school life.

Naturalistic and ethnographic studies. Naturalistic inquiry may refer either to a set of methods and perspectives that can enrich traditional evaluation research or to an alternative research paradigm that is fundamentally incompatable with any search for a single, objective reality (Guba, 1987). The comprehensive school evaluation envisioned in this paper would capitalize on naturalistic methods in the service of a traditional evaluation perspective. While recognizing the diversity of schools and the settings in which they function, school evaluations would

bring a common set of expectations and organizing principles to bear in judging a school's climate, effectiveness, and other qualities.

Naturalistic and ethnographic methods useful in school evaluation include relatively unstructured interviews, high-inference observations, and analyses of documents and records. Unobtrusive observations might be made of nonverbal cues as well as verbal behavior, especially in evaluating and documenting patterns of decision making; the school's climate; and other areas best communicated through intuitions, apprehensions, and other impressions, as well as propositional knowledge. These and related methods are discussed in Patton (1980), Miles and Huberman (1984), Goetz and LeCompte (1984), Lincoln and Guba (1985), and Fetterman (1988).

Educational connoisseurship and criticism. Eisner (1983) has described a distinctive approach to the study of classrooms, curriculum materials, and schools through educational connoisseurship and criticism. Connoisseurship is the art of appreciation, and criticism is the art of explanation or disclosure. The sensibilities of the connoisseur are formed and refined through a study of educatioanl theory, philosophy, and history to appreciate and evaluate educational activities. The connoisseur shares her perceptions through criticism, using metaphor and analogy as well as literal description to express the essence of the educational settings or materials studied. Criticism may be purely descriptive, or it may go beyond description to include interpretation and evaluation.

Quantitative descriptive studies. Sirotnik (1987) has proposed a school information system built around quantitative measures, which could be used to assess student learning outcomes, equity, and excellence, among other purposes. It would consist of an integrated database

incorporating data from students, teachers, administrators, and parents, as well as school records, and would include variables defined at the student, classroom, and school levels. Sirotnik's system would incorporate a range of student outcomes going well beyond those measured by test scores. It would support analyses of educational equity by showing whether amount and quality of educational resources were comparable across socioeconomic, racial/ethnic, and gender groups, and could address excellence by showing what proportion of students were achieving at the highest levels.

Variables Included

The variables used in quantitative descriptive studies are easy to characterize, but it is difficult to specify those employed in qualitative descriptive studies, because such studies rely more on text and narration than on numerical data and analysis. It may be useful, however, to consider some critical concepts or perspectives that help to assure the veracity and replicability of more qualitative descriptions.

Qualitative descriptive studies. In discussing the evaluation of accelerated schools for at-risk learners, Fetterman and Haertel (1989) discuss four concepts or perspectives that are basic to ethnographic approaches: intracultural diversity, contextualization, nonjudgmental orientation, and an emic perspective. The first two of these, intracultural diversity and contextualization, highlight the distinctiveness of schools, of the persons who inhabit them, and of the multiple classroom and other settings within a school. The last two concepts, a nonjudgmental orientation and an emic perspective, emphasize the importance of understanding the functioning of the school and the behavior of students

and staff in terms of their own perceptions, goals, and constraints. Taken together, these four principles lead the evaluator to seek multiple explanations for low achievement, dropping out, and other phenomena; aid in setting realistic expectations for the degree and rate of improvement possible; and above all highlight the complexity and uniqueness of schools as systems, and the futility of simplistic, top-down reform efforts. Useful, realistic analyses of a school's difficulties and prospects for improvement must be grounded in an understanding of the culture of that particular school.

Quantitative descriptive studies. The school information system proposed by Sirotnik (1987) begins with a student-level data base, including background data; attendance, suspensions and expulsions; grade point average, courses completed, track or program (academic, vocational, etc.), and special educational placements; performance on standardized as well as criterion-referenced tests; and assessments of higher-order thinking skills, oral and written communication, citizenship, and academic effort. These basic data would be supplemented with additional information on students' course taking and performance in school, their measured achievement, and their attitudes.

Additional data would be obtained from parents concerning home and family background, home learning environment and students' out-of-school activities, school-family relations, and parental perceptions of the school climate and learning environment. Teachers would contribute background data, and would be asked to report on their professional activities, attitudes, perceptions of the school and its leadership, and their educational philosophy and practices. Finally, the information system

envisoned by Sirotnik would employ classroom observations; interviews with teachers, students and principals; and analysis of documents to characterize each class of students within the school. Data at the student and class levels would be aggregated to the school level, and would be supplemented with school-level data on the overall schedule of course offerings, graduation requirements, and other variables that are defined most naturally at the school level.

Descriptive Studies and School Evaluation

By its nature, evaluation must go beyond description to include some judgment of worth or quality. For school evaluations, norm-referenced judgments of a school's processes and outcomes against those of comparable institutions have been widely used. Descriptive studies of schools rarely allow for such norm-referenced judgments. Naturalistic and ethnographic researchers often eschew judgments of worth, and although judgment is inherent in educational connoisseurship and criticism, it reflects the personal understanding of the writer. To the extent that variables could be defined consistently across schools, Sirotnik's (1987) more quantitative school information system could easily be extended to enable such comparisons. An alternative basis for judgment and evaluation, more in keeping with the particularistic focus of descriptive studies, is comparison to a school's own prior status or performance. Over time, as data are collected and records are maintained, change can be ascertained, at least permitting judgments of improvement or disimprovement.

Research using naturalistic methods also tends to require substantially more time and other resources than quantitative research.

For purposes of describing individual schools and analyzing their particular strengths and weaknesses, these costs may be acceptable, but for many evaluation purposes, the efficiencies of collecting and analyzing quantitative versus qualitative data will make quantitative methods more attractive.

Conclusions

Writing recently in the New York Times, Edward B. Fiske (1989) reported that "a movement is growing to grade schools, too, on classroom performance," and reported on a new initiative in New Jersey to inform parents annually of how their children's schools compare to other schools on such variables as reading scores, staffing ratios, and dropout rates. Fiske reported that such reports are already required in California, Illinois, and West Virginia, and there is interest in similar reports for the State of New York. The growth of indicator systems to promote school accountability, of school recognition programs, of commercial systems to implement the precepts of effective schools, and of other forms of school evaluation all attest to the increasing interest of researchers, policy makers, the public, and educators themselves in school evaluation.

Strengths and weaknesses of different approaches. The evaluation systems and approaches described in this paper were developed at different times, under different auspices, in response to different needs, and so it is not surprising that they vary considerably from one another. Despite their wide variability, each of the methods and approaches discussed has strengths as well as weaknesses, and taken together, they provide a useful point of departure in the attempt to formulate evaluation models.

Indicator and accountability systems have the advantage of drawing on data that are usually generated by schools, and thus by tradition have a good deal of face validity. But the linkages of input to output are often not clear, and variables representing schooling processes may be entirely absent from these systems. The variables included are too often limited to benchmark measures, in Oakes' (1986) terms, rather than variables that would show change.

The school recognition approach leads to a clear differentiation among schools with respect to outcomes, but again without a close linking of the input and process variables to those outcomes. Also, investigations of recognition criteria over time, and across grade levels and content areas within schools, raises serious questions about what is really being evaluated. Can we consider something so unstable to be a form of school evaluation?

The school effectiveness paradigm, coming from a research perspective, offers an empirical means of linking school practices with outcomes, but does not speak to the shortcomings of the outcome measures, nor to the weighting of input and process measures. For example, are the five effective schools factors equally important? Are they at least partially compensatory? Are they manipulable? If factors like effective leadership are not modifiable, then the effective schools model provides a description of status, but no guidance for improvement.

When the other approaches are appraised, a very different pattern of strengths and weaknesses arises. In the self-study approaches, there are a number of input and process measures that give a sense of what the school has been doing, and the possible trade-offs among measures. But

one is left with the question of whether it all adds up to a good school.

Accreditation follows a similar pattern, but adds the external judgment of experts (as does the USDE Recognition Program). This does offer a means of providing an objective judgment of school quality (arguably too low in accreditation and too high in recognition), but still relies on professional judgment that there is a linkage between the factors appraised and the outcomes presumed.

Finally, rich, contextualized school descriptions may offer a window on a school's culture and day-to-day life, and other variables believed to reflect what schooling is about. If this form of investigation leads to planful intervention and improvement, that is all to the good. But descriptive studies in themselves are not evaluations, and may stop short of yielding any judgment as to the merit or worth of the school.

Outcome-Oriented Versus Process-Oriented Evaluation Approaches

One major differentiation among the six approaches is between the three outcome-oriented methods of indicator and accountability systems, school recognition, and school effectiveness, and the three primarily process-oriented methods of self study, accreditation, and descriptive case studies. This distinction is reminiscent of the contrast between standardized test batteries and facilities audits, which have long existed side by side in schools. Schools are too complex to yield to either pole of such a dichotomy, however, and so it is our position that evaluation must encompass both. A comprehensive evaluation model must incorporate input, process, and outcome measures.

However, we need to acknowledge that the fundamental, epistemological differences reflected in these two general strategies will

not be solved simply by adding some variables from the other camp's artillery; indeed, this is already commonly done. Without major improvements in outcome measures and greatly increased knowledge of the ways that measured inputs and processes influence outcomes, the dichotomy will continue. Rather, we need to recognize that, given our present state of knowledge, each of these two strategies requires strong inferences from weak or unknown information, but they reflect very different decisions about where such inferences should occur. Those approaches that start from the outcome measures are willing to accept tenuous causal relationships back to inputs and processes, while the those starting from inputs and processes must be willing to trust that outcomes will follow.

Two implications for school-level evaluation follow from these observations. First, there needs to be agreement initially as to what type of inference is to be tolerated. There is little point in pursuing a heavily process-oriented evaluation where clients or audiences are interested only in an outcome-based one (Eichelberger, 1988).

More importantly, perhaps, those who conduct a particular type of evaluation are going to have to struggle, rationally and empirically, with justifying their inferences. Thus, those who start without a strong outcome link are going to have to build a strong argument that the school is achieving desired outcomes, whether by community attitudes, analogous results from similar schools, or other bases. It will also be wise for them to provide strong theoretical and logical arguments for the selection of input and process variables selected. Likewise, those who start from an outcome base must build a logical argument that the school's action could

reasonably be expected to be the cause of the results observed. They should also provide independent evidence (e.g., dropout rates, college enrollment of graduates, etc.) that the outcome is generalizable.

It also strikes us that the dichotomy in approaches may make less difference in actual school tests. It would be worthwhile to study whether different evaluation approaches reach different conclusions, and if so, what interactions are found between types of schools and evaluation approaches.

Decisions about the manner and extent to which each of the six evaluation approaches will be incorporated into an evaluation will reflect dimensions of variation in addition to the process-outcome distinction. Some of these have already been touched upon. These systems may also be contrasted in terms of (1) resource requirements, (2) degree of attention to the school context, (3) impetus for change and improvement at the school level versus a higher level of aggregation, (4) formative, criterion-referenced versus summative, norm-referenced evaluation focus, and reflecting all of the foregoing dimensions, (5) the categories of variables employed.

Resources Required

School evaluations require resources for data collection, analysis, and reporting. Typically the burden of providing data is shared by students, teachers, administrators, and sometimes observers from outside the system, although information collected for other purposes may also be used, as when school recognition programs screen potential applicants using data from ongoing testing programs. There are both advantages and disadvantages to having school personnel assume responsibility for data

collection and reporting. Such self-description may serve as the basis of a healthy process of self-study, but when the stakes are high, entirely internal evaluations may lack credibility.

Data analysis costs will depend largely on the methods of data collection employed and the level of detail at which results are to be reported. Clearly, analysis of naturalistic or ethnographic data can be far more costly than statistical analysis of quantitative data. The cost of preparing and disseminating evaluation reports will vary according to their degree of standardization, and according to the extent of dissemination. The per school cost of computer-generated school profiles may be low, but the cost of mailing them to all households with school children could be substantial. Reports required to rank schools or select reward recipients may be much simpler and less expensive to prepare than reports intended to diagnose problems or suggest improvement strategies for particular schools.

Contextualization

There is a tension between recognizing the unique context of each school versus comparing schools to one another according to a common set of criteria, or judging them against a common set of standards. Its resolution depends on the nature and purpose of the evaluation. In general, if the purpose is formative, seeking to clarify the functioning of a particular school and to guide its improvement, then standardization across schools is unimportant and contextualization is critical. A school's particular educational goals, strengths, and difficulties must bear on the interpretation of its processes and outcomes. If the purpose is primarily summative, ranking or rating schools according to global characteristics

so that rewards can be allocated or sanctions applied, then context, like outcomes, will probably be reduced to some small number of continuous variables, perhaps a single index of socioeconomic level. Interschool comparisons of any kind are likely to be unfair unless some adjustment is made for differences in socioeconomic level. It bears repeating, however, that such adjustments must never, even implicitly, serve to legitimate existing inequities in schooling processes or outcomes.

Locus of Decision Making and School Improvement

Policy makers often address educational problems as if the education were delivered through a rational, tightly coupled system in which goals established at state or district levels were faithfully translated into prescriptions for action at lower levels, and eventually implemented in the classroom. This is the conception implicit in many state accountability and indicator systems. An alternative view holds that schools are loosely coupled systems, in which change is best effected from within a single school, beginning with existing coalitions and interests to build consensus. These alternative perspectives will dictate the kinds of variables collected and analyses performed in a school evaluation. A view of schools as loosely coupled systems in which change must be bottom-up will indicate greater attention to the school's internal mechanisms of decision making; staff attitudes, values, and allegiances; and other process variables.

Criterion-Referenced Versus Norm-Referenced Comparisons

Evaluation implies some judgment of quality or value, usually through comparison to some standard. A school's processes and outcomes may be interpreted in the light of a fixed categories of educational quality or

defined standards for what is acceptable, or they may be interpreted via comparisons to other schools. Interpretations of the absolute levels of variables for a single school are criterion-referenced, and interpretations of a school's standing relative to other schools are norm-referenced. Of course, these two forms of comparison are not mutually exclusive. The appropriate form of comparison will depend on the purposes of the evaluation, and will in turn dictate the forms of data collected and the manner in which they are presented. For norm-referenced interpretations, comparability of measures across schools is paramount. Such interpretations are likely to be based on quantitative profiles of indicators. For criterion-referenced interpretations, the inherent meaningfulness of the data is of primary concern, and standardization acoss schools is less important. These requirements may be better met by narrative descriptions than numerical scores.

Variables Included

The purposes of an evaluation, its resource constraints and methods of data collection, its intended bases of comparison, and its treatment of the school context will all influence the categories of variables addressed. For purposes of school improvement, toward the formative end of the formative-summative continuum, softer methods and measures are likely to be more useful. For comparing schools to one another, toward the summative end of the continuum, more uniform, standardized, quantitative methods and measures will serve better.

Some possible variables are listed in an Appendix to this paper, under eight categories:

- the community and student population served by the school
- the school's physical plant and instructional facilities and resources
- the school's faculty, staff, and administration
- the school's philosophy and policies
- instructional processes, including provision for learners with special needs
- course offerings and overall program coordination
- cognitive learning outcomes
- other outcomes

Different categories may be more or less important, depending on the nature of the evaluation. Recall, for example, that outcome variables are only minimally represented in most school accreditation programs, whereas process variables are poorly represented in most accountability and indicator systems. At a minimum, a comprehensive school evaluation of any kind should probably include some variables representing context, instructional processes, and school outcomes.

Professional Standards for School Evaluation

From the foregoing discussion of different school evaluation models and of dimensions of variation among them, an argument can be made for greater attention to school-level evaluation as a specialized topic within evaluation theory and practice. Different evaluation approaches have confronted common problems, and all might profit from a pooling of good ideas and common solutions. The development of theory and improvement

of practice in school-level evaluation could be furthered significantly by the development of a set of professional standards for school evaluation.

The standards envisioned would complement those already developed for program evaluation (Joint Committee on Standards for Educational Evaluation, 1981), test use (AERA, APA, NCME, 1985; Committee on Fair Testing Practices, 1988) and personnel evaluation (Joint Committee on Standards for Educational Evaluation, 1988). They would not be narrowly prescriptive with regard to either content or methodology, but would serve to raise some of the kinds of questions that should be considered in connection with any school-level evaluation approach. Ideally, school evaluation standards would help to assure appropriate use of existing school evaluation models, encourage improvements to those models, and guide the evolution of new models to meet new needs.

The following are some suggestios of the issues that standards would have to address:

- The range of variables included. As discussed earlier, an evaluation must draw on input, process, and outcome variables. Further, there must be at least a defensible hypothesized relationship among them.
- Within-school variability. Studies and experience clearly indicate that the variability among classrooms and classes within schools is often much larger than that between schools. Designs that draw from a single class in a school, for example, are simply not defensible as school evaluations.
- Longitudinal trends. Broad inferences about school quality from data conducted at one point in time are likely to be seriously limited.

To give a notion of changes occurring in the school, and to verify cross-sectional findings, some type of longitudinal data collection is desirable.

Goodlad (1984, p. 31) attributed a lack of intelligent change to schools' lack of information about their own functioning. But from the foregoing review and discussion, it would appear that a lack of information per se is not the problem. The study of school evaluation as a specialization within the emergent discipline of evaluation, and the development of professional standards for school evaluation, would help to clarify the steps from knowledge accumulation to knowledge utilization.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Austin, G. R. (1979). Exemplary schools and the search for effectiveness. Educational Leadership, 37(1), 10-14.

Berman, P., & McLaughlin, M. W. (1977). Federal programs supporting educational change: Factors affecting implementation and continuation (Vol. 7). Santa Monica, CA: RAND Corporation.

Bryant, M. T. (1986). High school accreditation in California: A policy analysis of accreditation's role in the state initiated school reform movement. Unpublished doctoral disseration, Stanford University, Stanford, CA.

Buttram, J., Corcoran, T. B., & Hansen, B. J. (1986). Sizing up your school system: The district effectiveness audit. Trenton, NJ: New Jersey School Boards Association.

California State Department of Education. (1986). The California school recognition program. Sacramento, CA: Author.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). Equality of educational opportunity. Washington, DC: U. S. Government Printing Office.

Cuban, L. (1983). Effective schools: A friendly but cautionary note. Phi Delta Kappan, 64, 695-696.

Dade County Public Schools and United Teachers of Dade. (1986). QUIPP: The quality instruction incentives program. Miami: Author.

David, J. L. (1987). Improving education with locally developed indicators (CPRE Research Report Series RR-004). New Brunswick, NJ: Rutgers, the State University of New Jersey, Eagleton Institute of Politics, Center for Policy Research in Education.

DeRoche, E. F. (1987). An administrator's guide for evaluating programs and personnel (2nd ed.). Boston: Allyn and Bacon.

Dyer, H. S. (1972). Some thoughts about future studies. In F. Mosteller & D. Moynihan (Eds.), On equality of educational opportunity. New York: Vintage Press.

Edmonds, R. (1979). Effective schools for the urban poor. Educational Leadership, 37(1), 15-27.

Eichelberger, R. T. (1988, October). Assumptions made by non-positivistic evaluators: Do clients agree with them? Paper presented at the meeting of the American Evaluation Association, New Orleans.

Eisner, E. W. (1983). Educational connoisseurship and criticism: Their form and functions in educational evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models: Viewpoints on educational and human services evaluation (pp. 335-347). Boston, MA: Kluwer-Nijhoff.

Fetterman, D. M. (1988). Qualitative approaches to evaluating education. Educational Researcher, 17(8), 17-23.

Fetterman, D. M., & Haertel, E. H. (1989). A School-Based Evaluation Model for Accelerating the Education of Students At-Risk. Stanford, CA: Stanford University.

Fiske, E. B. (1989, March 1). Lessons. The New York Times, p. B6.

Goetz, J. P., & LeCompte, M. D. (1984). Ethnography and qualitative design in educational research. New York: Academic Press.

Goodlad, J. I. (1984). A place called school: Prospects for the future. New York: McGraw-Hill.

Grant, G. (1988). The world we created at Hamilton High. Cambridge, MA: Harvard University Press.

Guba, E. G. (1987). Naturalistic evaluation. In D. S. Cordray, H. S., Bloom, & R. J. Light (Eds.), New directions for program evaluation, No. 34: Evaluation practice in review (pp. 23-43). San Francisco, CA: Jossey-Bass.

Haertel, E. H. (1989). Within-state comparisons: Suitability of state models for national comparisons. In E. H. Haertel, et al., Report of the NAEP Technical Review Panel on the 1986 reading anomaly, the accuracy of NAEP trends, and issues raised by state-level NAEP comparisons (National Center for Education Statistics Report No. CS 89-499, pp. 229-250). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Joint Committee on Educational Evaluation. (1981). Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill.

Joint Committee on Educational Evaluation (1988). The personnel evaluation standards: How to assess systems for evaluating educators. Newbury Park, CA: Sage.

Klitgaard, R. E., & Hall, G. (1973). A statistical search for unusually effective schools. Santa Monica, CA: The RAND Corp.

Lightfoot, S. L. (1983). The good high school: Portraits of character and culture. New York: Basic Books.

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. Newbury Park, CA: Sage.

Mandeville, G. K. (1988). School effectiveness indices revisited: Cross-year stability. Journal of Educational Measurement, 25, 349-356.

Mandeville, G. K., & Anderson, L. W. (1987). The stability of school effectiveness indices across grade levels and subject areas. Journal of Educational Measurement, 24, 203-216.

May, J. (1987). An overview of the school incentive program. Columbia, SC: State Department of Education.

McLaughlin, M. W. (1978). Implementation as mutual adaptation: Change in classroom organization. In D. Mann (Ed.), Making change happen. New York: Teachers College Press.

Meyer, J. W., & Rowan, B. (1978). The structure of educational organizations. In M. Meyer, et al. (Eds.), Environments and organizations. San Francisco: Jossey-Bass.

Miles, M. B., & Huberman, A. M. (1984). Qualitative data analysis: A sourcebook of new methods. Newbury Park, CA: Sage.

Oakes, J. (1986). Educational indicators: A guide for policymakers (Center for Policy Research in Education [CPRE] Occasional Paper Series, Report No. OPE-01). Santa Monica, CA: The RAND Corporation. (Paper may be obtained from Publications Department, The RAND Corporation, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138)

OERI [Office of Educational Research and Improvement] State Accountability Study Group. (1988). Creating responsible and responsive accountability systems (Programs for the Improvement of Practice Report No. PIP 88-808). Washington, DC: United States Department of Education.

Patton, M. Q. (1980). Qualitative evaluation methods. Beverly Hills, CA: Sage.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. Elementary School Journal, 83, 427-452.

Ralph, J. H., & Fennessey, J. (1983). Science or reform: Some questions about the effective schools model. Phi Delta Kappan, 64, 689-694.

Scriven, M. (1987). Validity in personnel evaluation. Journal of Personnel Evaluation in Education, 1, 9-23.

Sirotnik, K. (1987). The information side of evaluation for local school improvement. International Journal of Educational Research, 11, 77-88.

Sizer, T. R. (1984). Horace's compromise: The dilemma of the American high school. Boston, MA: Houghton Mifflin.

Wynne, E. A. (Ed.) (1988). Designating winners: Using evaluation in school recognition programs (CSE Report No. 279). Los Angeles, CA: University of California, Graduate School of Education, Center for the Study of Evaluation.

Appendix

Example of Variables/Indicators for Use in Comprehensive School-Level Evaluation

1. Community and student population served by the school
   - 1.1. SES, demographics
   - 1.2. parent perceptions and attitudes
   - 1.3. student language background
   - 1.4. transiency, mobility
2. School physical plant, instructional facilities, and resources
   - 2.1. per pupil expenditures
   - 2.2. teacher salaries
   - 2.3. lighting, heating, ventilation, cleanliness, general upkeep
   - 2.4. special facilities--community outreach, transportation facilities permitting field trips
   - 2.5. learning labs, testing labs, mobile libraries, or other district-level facilities
   - 2.6. adequacy of school library
   - 2.7. computers, VCRs, instructional technology
   - 2.8. resources for teachers--zerox, telephones, etc.
3. Faculty, staff, and administration
   - 3.1. staffing pattern, teacher credentials, availability of specialists
   - 3.2. number of administrators, vice pricipals, curriculum coordinators
   - 3.3. amount of instructional time (time spent teaching) by faculty and staff
   - 3.4. pupil/teacher ratio
   - 3.5. number of aides--total picture of what's available for instruction
   - 3.6. teacher experience
   - 3.7. teachers' participation in continuing education
4. School philosophy and policies
   - 4.1. explicit homework policy
   - 4.2. attendance policy
   - 4.3. grading policy
   - 4.4. discipline policy
   - 4.5. guidelines for contacting parents
   - 4.6. school-wide achievement goals

4.7. written philosophy
4.8. alignment of instruction with school philosophy
4.9. agreement of individual staff with overall school philosophy
4.10. teacher inservice and staff development policies

5. Instructional processes
   5.1. within-classroom processes (e.g., Rosenshine's explicit teaching, variables examined by Berliner, classroom management, time on task, student engagement)
   5.2. classroom learning environments
   5.3. sensitivity to range in student abilities
   5.4. use of appropriate approaches for various children (e.g., adaptive learning)
   5.5. provision for learners with special needs
   5.6. use of cooperative learning and similar strategies
   5.7. peer tutoring

6. Course offerings and overall program coordination
   6.1. tracking or streaming
   6.2. coordination of pullout programs with regular classroom instruction
   6.3. use of flexible regrouping strategies, including cross-grade grouping
   6.4. curricular coherence for children with different abilities, interests, or needs

7. Cognitive learning outcomes
   7.1. state testing and assessment programs
   7.2. district-mandated standardized testing programs
   7.3. Coverage of cognitive content at each grade
   7.4. at high school level, need coverage of content areas
   7.5. AP course offerings, enrollments, and outcomes at the high school level
   7.6. student writing samples, portfolios, senior project ("capstone"), science fairs

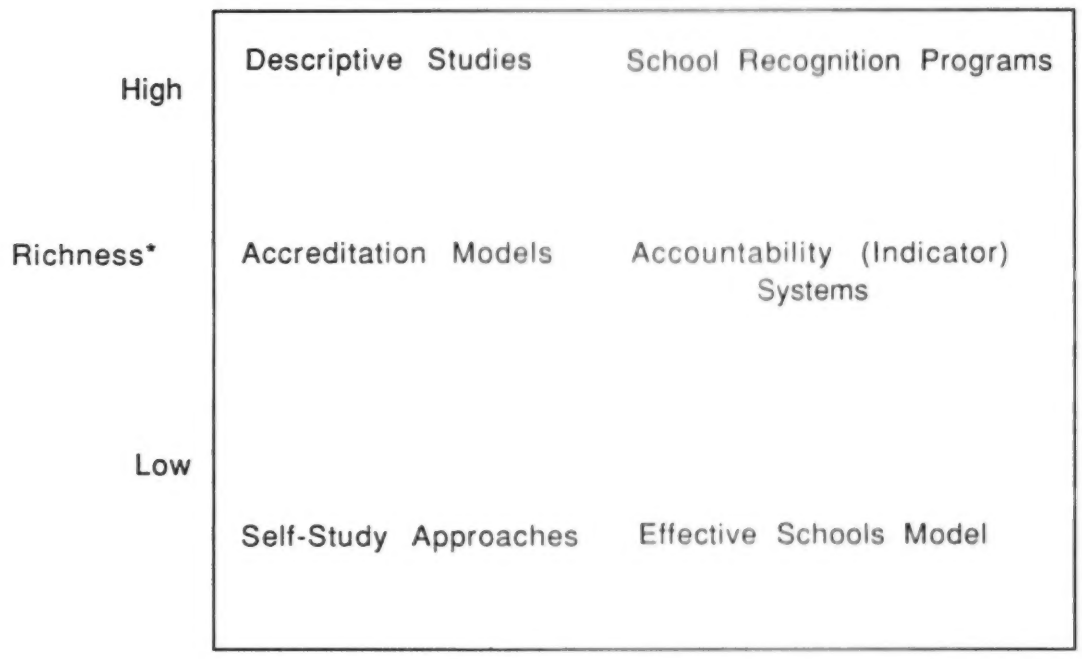
8. Other outcomes
   8.1. students' educational plans and expectations
   8.2. student attitudes toward school
   8.3. student attitudes toward subject matter areas
   8.4. student leisure reading and other outcomes of interest
   8.5. student dropout and attendance rates
   8.6. staff and teacher morale

8.7. teacher attendance
8.8. teacher attitudes toward principal and administration
8.9. teacher turnover
8.10. teacher perceptions of support, adequacy of materials, adequacy of compensation
8.11. safe school climate
8.12. parental attitudes, satisfaction with school
8.13. parental participation in school events
8.14. active PTA
8.15. community participation in school events

Figure Caption

Figure 1. Dimensions of variation among school evaluation approaches.

Candidate School Evaluation Models



*The vertical dimension, "richness," refers to the range of variables included in a model, the range of perspectives offered (teacher, parents, community, educational experts, administrators, etc.), and scientific credibility, including reliability and validity of the measures used.